Running title: Implementation of Physical Employment Standards

Abstract

Objective: The aim of this paper was to describe an approach to implementing and integrating physical employment standards into an organisational procedure, to ensure the safe and effective supervision of physical fitness of workers in a physically demanding occupation, using a real-world example. **Methods**: Using previously published cardiorespiratory, muscular strength and endurance physical demands data from UK firefighters, a process to manage all levels of physical capability was developed with industry stakeholders. **Results**: Performance standards and associated cut-scores relating to acceptable, uncertain, and unacceptable job performance, using a traffic-light style process, were agreed by stakeholders to ensure the safe and effective management of incumbent's physical fitness. **Conclusions**: This paper describes the processes involved in implementing a physical capability management procedure, for the administration of routine in-service physical employment standards and tests in the UK Fire & Rescue Service.

Key words:

Occupational Fitness; Physical Fitness; Fitness Management; Physical Employment Standards; Implementation

Introduction

Workers in physically demanding occupations, such as emergency or uniformed services, are often required to demonstrate appropriate levels of physical capability to undertake their role(s) safely and effectively. The physical attributes required for such professions are ordinarily assessed at the pre-employment (selection) stage to ensure that applicants can demonstrate the necessary physical attributes to undertake the work [1-4]. However, in some emergency service roles, concerns have been raised that the most demanding duties occur so infrequently that performing the job itself may be insufficient to maintain role-specific physical abilities [5]. Therefore, workers who fail to maintain appropriate levels of physical fitness from supplementary physical training throughout their career, put themselves at risk from over-exertion, possibly leading to injury or acute illness, which can be fatal [6-10]. This can also place work colleagues and the public at risk from failing to complete job tasks effectively in time-sensitive, emergency situations [11-14].

In recognition that physical fitness can impact the health, safety and operational performance of workers in physically arduous jobs, the implementation of robust, evidence-based physical employment standards (PES) to ensure both the initial and on-going physical competencies of workers have become increasingly important in recent years [12, 15-22]. However, the process of implementing PES that are valid, reliable [23-26], fairly applied and reasonable to all stakeholders [27, 28] can often be a challenging task for employers. Consequently, the implications of setting inappropriate standards can be costly to both the organisation and society, either through injury to employees or from applicants or incumbents being unfairly restricted from employment [12, 29, 30].

In response to these challenges, a united body of work by the scientific community has been established to standardise commonly used terms and phrases (presented in table 1) and to solidify a number of best-practice methods for the development of PES [1, 23, 24, 28-33].

[INSERT TABLE 1 ABOUT HERE]

However, despite the attention to this field of work, some of the more applied elements including the steps required to effectively integrate PES into organisational policies and procedures remain unclear. Indeed, little has been published articulating the most appropriate methods of safely managing incumbents that fail to meet PES due to a lack of physical ability or due to misclassifications in the testing process. In 2014, the American College of Occupational and Environmental Medicine identified that due to a lack of appropriate methodologies for fitness-for-duty assessments, industrial firefighters may not be being correctly assessed for their fitness for work [34]. These concerns were echoed in 2016, when Petersen et al. reported that there was an absence of resources to “advance knowledge and support best practice in this field” [29]. Specifically, there is a shortage of studies conveying the definitive step of describing how PES have been successfully integrated into organisational policies and procedures with the aim of managing all levels of physical capability in a safety-critical industry. To our knowledge, this will be the first paper to describe an approach used to integrate a developed PES with an associated management procedure, using a real-world example in the UK Fire & Rescue Service (UKFRS).

Methods

In 2012, the UKFRS established a collaboration between the Chief Fire Officers Association, the FireFit Steering Group and academics at the University of Bath to implement a research programme to investigate the cardiorespiratory, strength and muscular endurance demands of critical UK firefighting tasks and to identify minimum PES for safe and effective firefighting performance [19, 20, 22].

Project management

Prior to initiating the project, two distinct working groups were established to offer the research team with technical and strategic guidance relating to the job (e.g. UK firefighting) and to ensure senior management involvement. A Technical Panel (TP), consisting of operational subject matter experts (SME), was assembled to advise on the practical aspects of the job, whilst a Stakeholder Panel (SP) provided strategic direction to the project team and to ensure that the process and outcomes were both reasonable and justifiable to the customer.

Task analysis

The first phase of the research project was to conduct an up-to-date job task analysis of UK firefighting. A detailed 9-point process (modified from Tipton et al. 2012) was developed and outlined the specific steps required to: (1) establish the critical tasks; (2) determine the method of best practice of those tasks and; (3) agree on the minimum (acceptable) performance standards (MPS) for completing operational tasks for both firefighters (i.e. those involved in active firefighting duties) and incident commanders (i.e. those managing the operational incident) [30]. This was achieved by convening a series of workshops with the TP to follow the task analysis process and ultimately determine the minimum acceptable level of performance for each critical task. This was achieved using video analysis along with the Bookmark method of standards setting [35]. The TP were shown a video of each simulation of the critical task being performed at the three different paces (in sequence from slowest to fastest) with a detailed operational scenario being read out to them at the start of each video.

The panel were then asked to anonymously indicate on a scoring sheet the pace that they felt corresponded to the minimum acceptable performance of the specific task (within the context of the scenario described). For some tasks, such as lifting a mass overhead, successful or unsuccessful completion was discrete (pass/fail) and therefore did not require judgement on any appropriate pace. The individual votes from TP members were collated and presented back to the panel at the same meeting. The TP were then asked to reach a group consensus for each critical task through group discussion, had a consensus standard not already been identified. A comprehensive description of the task analysis process was published previously [22].

Physical demands analysis

Following the task analysis, two separate studies were conducted to investigate the cardiorespiratory demands [19] as well as the muscular strength and endurance demands of UK firefighting [20]. For the cardiorespiratory demands study, participants completed a number of standardised (critical) firefighting tasks (hose run (HR), equipment carry (EC), casualty evacuation (CE), stair climb (SC) and wildland fire (WF)) at a pre-determined MPS [22] to establish the peak steady-state metabolic cost of each task. Participants that were unable to maintain the MPS were deemed unsuccessful at completing the operational task and were removed from further analysis, as were the tasks that were considered unrealistic when compared to the 'actual job' [19]. The mean physical demand of participants that successfully completed the realistic firefighting tasks, i.e. those that maintained the MPS (CE, HR, EC, SC) were subsequently used to derive a minimum relative cardiorespiratory PES (i.e. maximum oxygen uptake; VO_2max in $ml.kg^{-1}.min^{-1}$) for use on generic predictive selection tests (PST) for both firefighting and incident command roles [19]. A comprehensive description of the physical demands analysis process for the determination of cardiorespiratory fitness is presented in more detail elsewhere [19].

For the muscular strength and endurance study, successful and unsuccessful completion of critical firefighting tasks, specifically two binary (pass/fail) ladder tasks (ladder lift and ladder lower), and one ladder extension task where participants were required to maintain the MPS, were compared with maximal strength and muscular endurance ability on three corresponding task related PST (seated shoulder press, seated single rope pull-down and seated repeated rope pull-down tests, respectively). These data were used to determine minimum strength and muscular endurance PES [20]. A comprehensive description the methodology used to develop muscular strength and endurance requirements are presented in more detail elsewhere [20].

Performance standards

For each PES, distinct levels of competence (i.e. performance standards) were described to clarify the proficiency at each specific level [36]. The performance standards were described as:

- Fail – A test score equivalent to unacceptable job performance (i.e. that is below the minimum level of physical capability for safe and effective work)
- Pass – A test score equivalent to acceptable job performance (i.e. that meets the minimum level of physical capability for safe and effective work)

Cut-score determination

For each PES, specific cut (passing) scores were established corresponding to acceptable job performance using a range of statistical methods:

Cardiorespiratory fitness standards

For the cardiorespiratory demands, the mean metabolic demand was calculated from all of the valid tasks and corrected for a realistically sustainable exercise intensity for the duration of the combination of tasks to reflect the physical demands of a generic emergency response [12, 37, 38]. These were calculated for both the firefighter role and incident command role as follows:

Firefighter cut-score – The mean metabolic cost for the four representative tasks (HR, CE, EC, SC) was 38.1 $ml.kg^{-1}.min^{-1}$ with the minimum expected duration of these tasks combined being 15:50 minutes. This length of task was deemed sustainable at 90% VO_2max [38] producing a resultant cut-score for cardiorespiratory fitness of 42.3 $ml.kg^{-1}.min^{-1}$ [19].

Incident commander cut-score – The only representative task for incident commanders was the SC task which had a metabolic cost of 34.7 $ml.kg^{-1}.min^{-1}$. With a task duration of 06:04 minutes, a sustainable work intensity of 95% VO_2max was deemed appropriate producing a cut-score of 36.8 $ml.kg^{-1}.min^{-1}$ [19].

Muscular strength and endurance standards

For the strength and muscular endurance tasks, cut-scores associated with acceptable job performance (i.e. the MPS) were established by determining the most optimum balance of test sensitivity and specificity whilst maintaining test specificity of 90% or greater. This was achieved using contingency tables along with receiver-operating characteristic (ROC) curves and created cut-scores of 35kg for the seated shoulder press test (specificity 100%), 60kg for the seated single rope pull-down test (specificity 92%) and 23 repetitions of the 28kg weight in the seated repeated rope pull-down test (specificity 93%) [20]. A comprehensive description of the methods used in the physical demands analysis process to determine cut-scores are presented elsewhere for both the cardiorespiratory fitness standards [19] and strength and muscular endurance fitness standards [20].

Cut-score uncertainty

On establishing the cut-scores for each performance standard, it became clear that for each PES a 'zone of uncertainty', as described by Petersen et al. (2016), existed below the pass score where both true negatives and false negative results were present. The performance standards were subsequently amended to recognise this group, as follows:

- Fail – A test score equivalent to unacceptable job performance (i.e. that is below the minimum level of physical capability for safe and effective work)
- Unclear – A test score equivalent to uncertain job performance
- Pass – A test score equivalent to acceptable job performance (i.e. that meets the minimum level of physical capability for safe and effective work)

Following this, cut-scores relating to unacceptable job performance for each of the performance standards were calculated, thus creating an 'uncertain' zone between acceptable and unacceptable performance. Unfortunately, there is no clear, recognised best-practice method for how these zones should be defined in the PES literature. Depending on the study design and the type of test, standard or parameters being measured, these boundaries could be based on a variety of methods. These include (but are not limited to), the expected variance in the workforce indicated by the sample population, the error or reliability statistics of the predictive test or the level of variance or error observed while developing the PES. In this case, the cardiorespiratory fitness, muscular strength and endurance standards were treated differently due to the metabolic demands data being continuous and the strength and muscular endurance demands data being discrete relating to the standard weight increments used in PST equipment (seated shoulder press, seated single rope pull-down and seated repeated rope pull-down tests).

For cardiorespiratory fitness, the cut-score for unacceptable job performance was calculated by subtracting the mean standard deviation for the valid tasks from the MPS creating a VO_2max cut-score of 35.6 $ml.kg^{-1}.min^{-1}$ (42.3-6.7 $ml.kg^{-1}.min^{-1}$) for those in firefighting roles and 31.4 $ml.kg^{-1}.min^{-1}$ (36.8-5.4 $ml.kg^{-1}.min^{-1}$) for those in incident command roles. For strength and muscular endurance, the cut-score for unacceptable job performance was determined by identifying a point on the test score scale below the MPS with a test sensitivity as close to but not less than 90%, creating cut-scores of 32.5 kg for the seated shoulder press test (sensitivity 100%), 52 kg for the seated single-rope pull-down test (sensitivity 91%) and 15 repetitions of a 28 kg weight for the seated repeated rope pull-down test (sensitivity 90%).

Management process

Through discussion with both the TP and SP, it was suggested that the performance standards (fail/unclear/pass), could be colour coded to resemble a modified traffic-light system or RAG (Red, Amber, Green) rating. This would allow all stakeholders in the organisation involved in managing the physical capability of incumbents (e.g. health and fitness advisors, occupational health clinicians and human resource managers) to easily understand how physical fitness relates to occupational performance without the need to understand the scientific discipline(s) which are used to derive PES. The performance standards and associated management categories (colours) were proposed as:

- Fail - red

- Unclear - amber
- Pass - green

Through further consultation with the TP and SP, it was suggested that the management procedure should include a standardised process to manage all employees entering into routine physical capability testing (e.g. health screening). Additionally, an agreed plan should be in place to support incumbents to improve physical fitness, should they fail any of the physical capability tests. It was also suggested that tests should be conducted at least once a year for all operational personnel to ensure the maintenance of physical fitness.

Results

Management process

The traffic-light style process for managing physical capability was agreed through consultation with both the TP and SP. This process involved the recommendation that incumbents undertake a battery of PST once a year (i.e. VO_2max, shoulder press, single rope pull-down and repeated rope pull-down tests) following a recognised pre-exercise health screening process [39, 40]. Based on the PST results, incumbents were categorised as either fail (red), unclear (amber), pass (green) for each test. Those with all test scores in the 'pass' category were deemed physically capable for operational duties and no further action was necessary. Incumbents with a test score in the 'fail' category for any of the PST were deemed physically incapable and were recommended to be temporarily removed from operational duties. It was suggested that personnel should undergo physical training to improve specific fitness levels and pass a retest prior to returning to full operational duties. Those with any test score in the 'unclear' category were deemed to have uncertain physical fitness and were subsequently required to undertake either a timed direct task simulation (DTS) (for cardiorespiratory fitness) or criterion task (for strength and muscular endurance) to clarify their physical capability i.e. either physically capable (green) or incapable (red). This process is shown in Figure 1 below.

Figure 1. The generic fitness management process for the UKFRS

[INSERT FIGURE 1 ABOUT HERE]

Cardiorespiratory fitness

Specific cut-scores for each of the performance standards are identified in Tables 2 and 3. For cardiorespiratory fitness (Table 2), individuals in a firefighting role with a relative VO_2max of 42.3 $ml.kg^{-1}.min^{-1}$ or greater were considered physically capable whilst those with a VO_2max of 35.5 $ml.kg^{-1}.min^{-1}$ or less were considered incapable for operational duties. Those in-between (35.6-42.2 $ml.kg^{-1}.min^{-1}$) were subsequently required to complete a DTS with a pass time equivalent to the MPS of 42.3 $ml.kg^{-1}.min^{-1}$ (on successful completion of a high intensity physical activity screen), to ultimately ascertain their physical capability (i.e. capable or incapable) to undertake safe and effective work [18].

Individuals in an incident command role with a relative VO_2max of 36.8 $ml.kg^{-1}.min^{-1}$ or greater were considered physically capable, whilst those with a VO_2max of 31.3 $ml.kg^{-1}.min^{-1}$ or less were considered incapable for operational duties. Those in-between (31.4-36.7 $ml.kg^{-1}.min^{-1}$) triggering a medical assessment of cardiovascular risk factors [41] to determine their medical risk for moderate to high intensity physical activity. Individuals in this ‘uncertain’ category with up to 1 risk factor were considered physically capable for operational duties, whilst those with 2 or more were considered incapable. This alternative process for managing incident commanders is related to the lower overall cardiorespiratory fitness requirements for this role and the lack of a reproducible criterion task for this group.

[INSERT TABLE 2 ABOUT HERE]

Strength and muscular endurance

For strength and muscular endurance (table 3), incumbents that were able to shoulder press 35 kg or greater from an upright seated position on the shoulder press test were considered physically capable, whilst those that could only lift 30 kg or less were considered incapable of conducting safe and effective work. For the ladder lower task, firefighters that could pull down 60 kg or more on a single seated rope pull-down test were considered physically capable whilst those that could only pull down 51 kg or less were considered incapable. Finally, on the ladder extension task, firefighters that could complete more than 23 repetitions on a seated 28 kg repeated rope-pull down task were considered physically capable whilst those that could only complete 14 repetitions or less were considered incapable. For all three strength and muscular endurance tests, performances in the 'unclear' category triggered the requirement to undertake the relevant criterion assessment using a simulator for the ladder lift task [2] or standard operational equipment for the ladder lower and ladder extension tasks [20] to ascertain their physical capability (i.e. capable or incapable) to undertake safe and effective work.

[INSTER TABLE 3 ABOUT HERE]

Discussion

The implementation of PES into routine physical capability testing for organisations involving physically demanding work has become increasingly important in recent years with the acknowledgment that employers have a duty of care to ensure the on-going physical capability of its employees [19, 29, 42]. This is in part due to the greater understanding of how physical fitness interacts with operational capability and the risks associated with physically demanding emergency response work [6-8, 10, 43]. However, despite decades of investigation into PES methodology, few resources are available to promote best practice in this field with little attention being focused on the approaches used to successfully integrate PES into organisational policies and procedures [29, 34]. This is surprising considering the importance of routine physical capability testing (particularly for emergency service workers), to ensure the safety of incumbents, work colleagues and the public.

The purpose of this paper, therefore, was to describe the processes involved in integrating a physical capability testing process into organisational procedures for the administration of routine in-service PES in a physically demanding occupation. This was achieved by (1) defining performance standards , (2) identifying cut-scores associated with each performance standard using physical demands analysis data [19, 20] and (3) agreeing a process with industry stakeholders for triaging and managing test performances [18]. This was represented using a modified traffic-light system as ‘red’ (physically incapable); ‘amber’ (unclear) and ‘green’ (physically capable) to ensure the PES were easy to understand to all stakeholders in the organisation.

Implementing an organisation-wide, routine physical capability testing process can be onerous for employers and often involves both theoretical and practical challenges to overcome to ensure PES are properly administered. Some of the more theoretical considerations may often come in the development stage of establishing a PES, including defining role related tasks [44, 45], determining minimum performance standard(s) [1, 12] or determining appropriate cut-scores for PES [24, 36]. The practical issues may well often follow and be dominated by factors such as the resources and logistics needed to implement a service-wide testing programme for hundreds, or even thousands of employees spread across multiple work sites. Other practical factors may include finances for testing equipment, the time taken to administer each battery of tests as well as the knock-on effect to operational availability, which will all influence the decisions around how PES are implemented. The physical capability tests used in conjunction with PES are often debated and whilst both PST and DTS are regularly used for the assessment of both applicants and incumbents [1, 2, 19, 20, 46-49], limitations have been identified with both approaches when used for the assessment of individual's appropriateness for work [1, 18, 30, 32, 50-52].

Whichever the type of test used, it remains important to demonstrate they are valid, reliable and ultimately differentiate physically capable from incapable workers [23]. However, whilst the theoretical and often traditional approach to standard-setting attempts to delineate acceptable from unacceptable job performance with two distinct (i.e. pass/fail) performance standards, this approach rarely differentiates perfectly those that can from those that cannot perform the job [36]. Consequently, this can lead to incorrectly classifying individuals as physically capable (false positive result) or incapable (false negative result) with the consequences potentially leading to serious injury or unfair termination of employment, respectively. Whilst it is recognised that a degree of uncertainty exists around every cut-score [29], the decision on where to set the cut-point (passing score) will ultimately influence the safety of workers and/or their human rights depending where on the test-score scale it is set [24].

The issue of dealing with false positive and false negative results can therefore be remarkably challenging. Whilst an organisation in their duty of care (particularly those in the emergency services), may wish to reduce the number of false positive results thus not putting potentially unfit workers into unsafe situations, this will more often than not increase the number of fit incumbents being unfairly removed from work and possibly unfairly terminated from employment [28]. This may be unacceptable to stakeholders such as trade unions. Conversely, reducing the number of false negative results, would subsequently increase the risk to the organisation of allowing more unfit workers to remain on operational duties thus potentially increasing the risk of injury, illness or worse.

With the acknowledgment that the standard-setting process often contains inherent difficulties, organisations must develop a process of managing test scores that fall within this category of test variability, which may not have been routinely considered in the past [29]. This requires management, trade unions and occupational scientists to work together to determine PES, tests and processes that are both practical and scientifically valid but at the same time reasonable to all stakeholders. Such a framework could help to minimise test misclassifications from factors such as test validity when using PST [50, 53] and test familiarisation [54] or biological variability [51] when using DTS, suggesting that an updated approach may be warranted to ensure that PES are more accurately implemented [29]. Indeed, moving away from the binary pass/fail performance standards to a three standard (pass/unclear/fail) approach with a process for supplementary testing to clarify fitness for duty may create a means to satisfy all stakeholders and therefore more successfully integrate PES into organisational policies and procedures.

In the present study, a mix of PST, DTS and defined job tasks were used in an attempt to deliver an effective testing procedure for the UKFRS whilst ensuring the human rights and safety of emergency service workers [18-20, 55]. Following the analysis of the physical demands data, it became clear that the adoption of three performance standards (and associated cut-scores) relating to unacceptable, uncertain, and acceptable job performance with an associated management procedure were necessary in order to categorise employees appropriately and to minimise misclassifications in the testing process. A range of statistical methods were also required to determine appropriate cut-scores for each of the physical employment standards.

For the cardiorespiratory fitness standards, cut-scores relating to acceptable job performance (i.e. the MPS) were determined by calculating the mean metabolic demand from the valid tasks (and corrected for a realistically sustainable exercise intensity for the duration of the combined tasks) to reflect the physical demands of a generic emergency response, a method previously reported by Bilzon et al. [12]. The unacceptable job performance standard and subsequent 'unclear' zone for firefighter cardiorespiratory fitness was determined by subtracting the mean standard deviation from the critical tasks from the MPS. [19]. This accommodated individuals that were able to successfully complete firefighting tasks with a lower VO_2max than the derived MPS by providing them with the opportunity of undertaking a DTS to demonstrate their fitness for work (i.e. minimising false negative test results). These calculations were possible due to the collection and subsequent analysis of continuous metabolic (VO_2) data.

For the muscular strength and endurance PES cut -scores were determined using contingency tables along with receiver-operating characteristic curves. This approach was adopted as the strength and muscular endurance data was discrete due to the standard integers (i.e. 2.5kg increments) commonly found when using resistance equipment for PST. Acceptable job performance standards (i.e. the MPS) were established by determining the most optimum balance of test sensitivity and specificity whilst maintaining test specificity of 90% or greater. The unacceptable job performance standard and subsequent 'unclear' zones were calculated by identifying test scores below the MPS as close to (but not lower) than 90% test sensitivity. This approach when used in conjunction with the triage process ensured that both false positive and false negative results remained low (i.e. less than 10%). When test sensitivity and specificity are known to be in direct opposition, this can often be challenging to achieve. Therefore, developing an 'unclear' zone with further (clarification) testing, may be considered a reasonable approach to satisfy both the health and safety and human rights concerns of policy makers when developing PES, particularly those involved in safety critical emergency-service work.

Finally, the PES and physical capability tests were brought together into a simple, visual, easy to understand traffic-light style management process for the integration into organisational procedures. This ensured that incumbents who failed a fitness test were removed from operational duties, thus ensuring the health and safety of employees and the public. These employees were referred to a service health and fitness adviser for fitness training support to assist the employee until their fitness had improved to the requisite level. Incumbents with an 'unclear' test score were (assuming they were medically safe to do so) given an opportunity to undertake a further test to demonstrate their physical ability to undertake the operational work in question, thus minimising unfairly and unnecessarily removing an employee from duty. Due to a lack of published guidance on the ways to develop 'unclear' cut-scores, this paper was not able to follow any best-practice model for developing these procedural steps for integration into the UKFRS. However, through working closely with management, trade unions, subject-matter experts and stakeholders, a process was developed which serves as a good starting point in moving this area of PES development forward.

In summary, this paper describes the processes involved in implementing a physical capability management procedure, for the administration of routine in-service PES and tests in a physically demanding occupation and was achieved by defining performance standards, identifying cut-scores associated and agreeing a process with industry stakeholders for triaging and managing test performances. This process, developed in partnership with the UKFRS, trade union representatives and relevant government departments, could be applied to other public safety occupations to ensure the safe and effective management of employee physical fitness.

References

1. Jamnik, V., R. Gumienak, and N. Gledhill, *Developing legally defensible physiological employment standards for prominent physically demanding public safety occupations: a Canadian perspective.* Eur J Appl Physiol, 2013. **113**(10): p. 2447-2457.
2. Blacker, S.D., M.P. Rayson, D.M. Wilkinson, et al., *Physical employment standards for U.K. fire and rescue service personnel.* Occup Med (Lond), 2016. **66**(1): p. 38-45.
3. Williams-Bell, F.M., R. Villar, M.T. Sharratt, and R.L. Hughson, *Physiological demands of the firefighter Candidate Physical Ability Test.* Med Sci Sports Exerc, 2009. **41**(3): p. 653-662.
4. Gebhardt, D.L., *Historical perspective on physical employment standards.* Work, 2019: p. 1-14.
5. Rayson, M.P., D.M. Wilkinson, J.M. Carter, and A.M. Nevill, *National Firefighter Selection Process Development and Validation of National Firefighter Selection Tests: Physical Tests.* 2009a: London, UK: Department for Communities and Local Government.
6. Baur, D.M., C.A. Christophi, A.J. Tsismenakis, E.F. Cook, and S.N. Kales, *Cardiorespiratory fitness predicts cardiovascular risk profiles in career firefighters.* J Occup Environ Med, 2011. **53**(10): p. 1155-1160.
7. Durand, G., A.J. Tsismenakis, S.A. Jahnke, et al., *Firefighters' physical activity: relation to fitness and cardiovascular disease risk.* Med Sci Sports Exerc, 2011. **43**(9): p. 1752-1759.
8. Kales, S.N., E.S. Soteriades, C.A. Christophi, and D.C. Christiani, *Emergency duties and deaths from heart disease among firefighters in the United States.* N Engl J Med, 2007. **356**(12): p. 1207-1215.

9. Smith, D.L., D.A. Barr, and S.N. Kales, *Extreme sacrifice: sudden cardiac death in the US Fire Service.* Extrem Physiol Med, 2013. **2**(1): p. 6.

10. Baur, D.M., A. Leiba, C.A. Christophi, and S.N. Kales, *Low fitness is associated with exercise abnormalities among asymptomatic firefighters.* Occup Med (Lond), 2012. **62**(7): p. 566-569.

11. Elsner, K.L. and F.W. Kolkhorst, *Metabolic demands of simulated firefighting tasks.* Ergonomics, 2008. **51**(9): p. 1418-1425.

12. Bilzon, J.L., E.G. Scarpello, C.V. Smith, N.A. Ravenhill, and M.P. Rayson, *Characterization of the metabolic demands of simulated shipboard Royal Navy fire-fighting tasks.* Ergonomics, 2001. **44**(8): p. 766-780.

13. Holmer, I. and D. Gavhed, *Classification of metabolic and respiratory demands in fire fighting activity with extreme workloads.* Appl Ergon, 2007. **38**(1): p. 45-52.

14. Wynn, P. and P. Hawdon, *Cardiorespiratory fitness selection standard and occupational outcomes in trainee firefighters.* Occup Med (Lond), 2012. **62**(2): p. 123-128.

15. Dreger, R.W. and S.R. Petersen, *Oxygen cost of the CF-DND fire fit test in males and females.* Appl Physiol Nutr Metab, 2007. **32**(3): p. 454-462.

16. Plat, M.J., M.H. Frings-Dresen, and J.K. Sluiter, *Reproducibility and validity of the stair-climb test for fire fighters.* Int Arch Occup Environ Health, 2010. **83**(7): p. 725-731.

17. Todd Rogers, W., D. Docherty, and S. Petersen, *Establishment of performance standards and a cut-score for the Canadian Forces firefighter physical fitness maintenance evaluation (FF PFME).* Ergonomics, 2014. **57**(11): p. 1750-1759.

18. Stevenson, R.D.M., A.G. Siddall, P.J.F. Turner, and J.L.J. Bilzon, *Validity and Reliability of Firefighting Simulation Test Performance.* J Occup Environ Med, 2019. **61**(6): p. 479-483.

19. Siddall, A.G., R.D. Stevenson, P.F. Turner, K.A. Stokes, and J.L. Bilzon, *Development of role-related minimum cardiorespiratory fitness standards for firefighters and commanders.* Ergonomics, 2016. **59**(10): p. 1335-1343.

20. Stevenson, R.D., A.G. Siddall, P.F. Turner, and J.L. Bilzon, *Physical Employment Standards for UK Firefighters: Minimum Muscular Strength and Endurance Requirements.* J Occup Environ Med, 2017. **59**(1): p. 74-79.

21. Kenny, G.P., H. Groeller, R. McGinn, and A.D. Flouris, *Age, human performance, and physical employment standards.* Appl Physiol Nutr Metab, 2016. **41**(6 Suppl 2): p. S92-S107.

22. Stevenson, R.D., A.G. Siddall, P.F. Turner, and J.L. Bilzon, *A Task Analysis Methodology for the Development of Minimum Physical Employment Standards.* J Occup Environ Med, 2016. **58**(8): p. 846-851.

23. Milligan, G.S., T.J. Reilly, B.D. Zumbo, and M.J. Tipton, *Validity and reliability of physical employment standards.* Appl Physiol Nutr Metab, 2016. **41**(6 Suppl 2): p. S83-91.

24. Payne, W. and J. Harvey, *A framework for the design and development of physical employment tests and standards.* Ergonomics, 2010. **53**(7): p. 858-871.

25. Innes, E. and L. Straker, *Reliability of work-related assessments.* Work, 1999. **13**(2): p. 107-124.

26. Innes, E. and L. Straker, *Validity of work-related assessments.* Work, 1999. **13**(2): p. 125-152.

27. Great Britain Parliament, *Equality Act.* London: HSMO, 2010.

28. Adams, E.M., *Human rights at work: Physical standards for employment and human rights law.* Appl Physiol Nutr Metab, 2016. **41**(6 Suppl 2): p. S63-73.

29. Petersen, S.R., G.S. Anderson, M.J. Tipton, et al., *Towards best practice in physical and physiological employment standards.* Appl Physiol Nutr Metab, 2016. **41**(6 Suppl 2): p. S47-62.

30. Tipton, M.J., G.S. Milligan, and T.J. Reilly, *Physiological employment standards I. Occupational fitness standards: objectively subjective?* Eur J Appl Physiol, 2013. **113**(10): p. 2435-2446.

31. Milligan, G.S., S.D. Blacker, P.E.H. Brown, and A.G. Siddall, *The Third International Conference on Physical Employment Standards.* Work, 2019. **63**(4): p. 479-480.

32. Rogers, T.W., D. Docherty, and S. Petersen, *Establishment of performance standards and a cut-score for the Canadian Forces Firefighter Physical Fitness Maintenance Evaluation (FF PFME).* Ergonomics, 2014. **57**(11): p. 1750-1759.

33. Blacklock, R.E., T.J. Reilly, M. Spivock, P.S. Newton, and S.M. Olinek, *Standard Establishment Through Scenarios (SETS): A new technique for occupational fitness standards.* Work, 2015. **52**(2): p. 375-383.

34. Bhojani, F.A., L.A. Castillejo-Picco, D. Cathcart, et al., *Fitness-for-Duty Assessments of Industrial Firefighters.* Journal of Occupational and Environmental Medicine, 2018. **60**(2): p. e82-e89.

35. Lewis, D.M., Mitzel, H. C., Green, D. R., & Patz, R. J., *The Bookmark standard setting procedure.* Monterey, CA: McGraw-Hill, 1999.

36. Zumbo, B.D., *Standard-setting methodology: Establishing performance standards and setting cut-scores to assist score interpretation.* Appl Physiol Nutr Metab, 2016. **41**(6 Suppl 2): p. S74-82.

37. Blondel, N., S. Berthoin, V. Billat, and G. Lensel, *Relationship between run times to exhaustion at 90, 100, 120, and 140% of vVO2max and velocity expressed relatively to critical velocity and maximal velocity.* Int J Sports Med, 2001. **22**(1): p. 27-33.

38. Louhevaara, V., J. Smolander, O. Korhonen, and T. Tuomi, *Maximal working times with a self-contained breathing apparatus.* Ergonomics, 1986. **29**(1): p. 77-85.

39. Bredin, S.S.D., N. Gledhill, V.K. Jamnik, and D.E.R. Warburton, *PAR-Q plus and ePARmed-X plus New risk stratification and physical activity clearance strategy for physicians and patients alike.* Canadian Family Physician, 2013. **59**(3): p. 273-277.

40. Green, M., *Risk Stratification: Effective Use of ACSM Guidelines and Integration of Professional Judgment.* Acsms Health & Fitness Journal, 2010. **14**(4): p. 22-28.

41. American College of Sports Medicine, *ACSM's Resource Manual for Guidelines for Exercise Testing and Prescription. 6th ed.* 2010.

42. Health and Safety at Work Act, *Elizabeth II. Chapter 22. (1974) London: The Stationary Office. Available at: http://www.legislation.gov.uk/ukpga/1974/37/contents (Accessed 22nd November 2016)*. 1974.

43. Baur, D.M., C.A. Christophi, and S.N. Kales, *Metabolic syndrome is inversely related to cardiorespiratory fitness in male career firefighters.* J Strength Cond Res, 2012. **26**(9): p. 2331-2337.

44. Rue, C.A., M.P. Rayson, E.F. Walker, et al., *A job task analysis to describe the physical demands of specialist paramedic roles in the National Ambulance Resilience Unit (NARU).* Work, 2019. **63**(4): p. 547-557.

45. Taylor, N.A., H.H. Fullagar, B.J. Mott, J.A. Sampson, and H. Groeller, *Employment Standards for Australian Urban Firefighters: Part 1: The Essential, Physically Demanding Tasks.* J Occup Environ Med, 2015. **57**(10): p. 1063-1071.

46. Lindberg, A.S., J. Oksa, and C. Malm, *Laboratory or field tests for evaluating firefighters' work capacity?* PLoS One, 2014. **9**(3): p. e91215.

47. Lindberg, A.S., J. Oksa, D. Gavhed, and C. Malm, *Field tests for evaluating the aerobic work capacity of firefighters.* PLoS One, 2013. **8**(7): p. e68047.

48. Larsson, H., M. Tegern, A. Monnier, et al., *Content Validity Index and Intra- and Inter-Rater Reliability of a New Muscle Strength/Endurance Test Battery for Swedish Soldiers.* PLoS One, 2015. **10**(7): p. e0132185.

49. Foulis, S.A., M.A. Sharp, J.E. Redmond, et al., *U.S. Army Physical Demands Study: Development of the Occupational Physical Assessment Test for Combat Arms soldiers.* J Sci Med Sport, 2017. **20 Suppl 4**: p. S74-S78.

50. Buckley, J.P., J. Sim, R.G. Eston, R. Hession, and R. Fox, *Reliability and validity of measures taken during the Chester step test to predict aerobic power and to prescribe aerobic exercise.* Br J Sports Med, 2004. **38**(2): p. 197-205.

51. Boyd, L., T. Rogers, D. Docherty, and S. Petersen, *Variability in performance on a work simulation test of physical fitness for firefighters.* Appl Physiol Nutr Metab, 2015. **40**(4): p. 364-370.

52. Jamnik, V.K., S.G. Thomas, and N. Gledhill, *Applying the Meiorin Decision requirements to the fitness test for correctional officer applicants; examining adverse impact and accommodation.* Appl Physiol Nutr Metab, 2010. **35**(1): p. 71-81.

53. Morris, M., E. Deery, and K. Sykes, *Chester treadmill police tests as alternatives to 15-m shuttle running.* Occup Med (Lond), 2019. **69**(2): p. 133-138.

54. Fullagar, H.H., J.A. Sampson, B.J. Mott, et al., *Employment Standards for Australian Urban Firefighters: Part 4: Physical Aptitude Tests and Standards.* J Occup Environ Med, 2015. **57**(10): p. 1092-1097.

55. Siddall, A.G., R.D.M. Stevenson, P.J.F. Turner, and J.L.J. Bilzon, *Physical and Physiological Performance Determinants of a Firefighting Simulation Test.* J Occup Environ Med, 2018. **60**(7): p. 637-643.